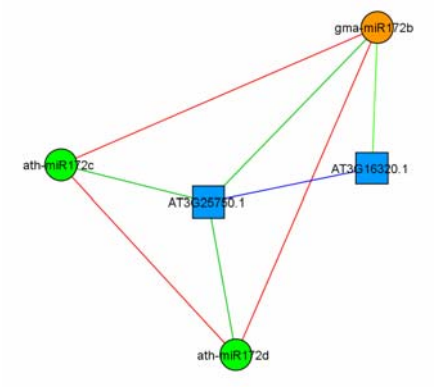
CrossLink User Manual

by Tobias Dezulian
(v1.2.4) 31.03.2006 11:04:22



Abstract 2
Usage overview 3
Quick beginner's tour 4
User Interface 10
  Control Panel 10
  Visualization Window 14
Example templates 14
Terms and Conditions 16

This manual adds detail information to the CrossLink article which is available from this website:
http://www-ab.informatik.uni-tuebingen.de/software/crosslink/doc/welcome.html

Please read the the CrossLink article first.

Abstract

CrossLink is a versatile tool for the exploration of relationships between RNA sequences. After a parametrization phase, CrossLink delegates the determination of sequence relationships to established tools (BLAST, Vmatch and RNAhybrid) and then constructs a network. Each node in this network represents a sequence and each link represents a match or a set of matches. Match attributes are reflected by graphical attributes of the links and corresponding alignments are displayed on a mouse-click. The distributions of match attributes such as E-value, match length and proportion of identical nucleotides are displayed as histograms. Sequence sets can be highlighted and visibility of designated matches can be suppressed by real-time adjustable thresholds for attribute combinations. Powerful network layout operations (such as spring-embedding algorithms) and navigation capabilities complete the exploration features of this tool.

CrossLink can be especially useful in a microRNA context since Vmatch and RNAhybrid are suitable tools for determining the antisense and hybridisation relationships which are decisive for the interaction between microRNAs and their targets.

CrossLink is available both online and as a standalone version at http://www ab.informatik.uni tuebingen.de/software.

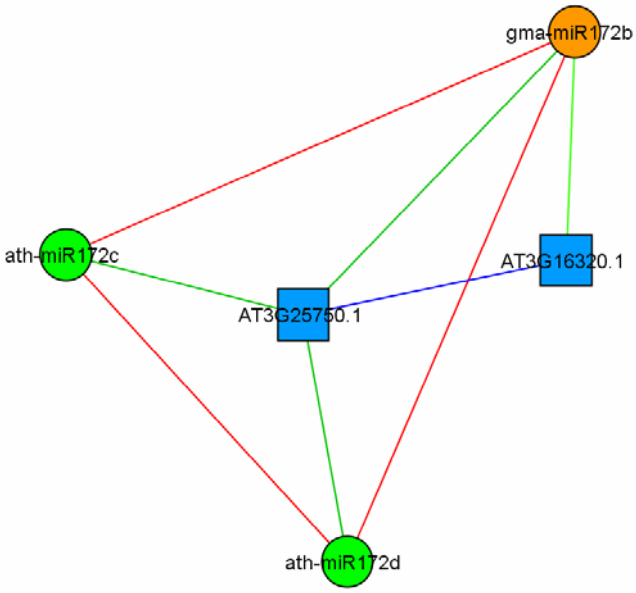
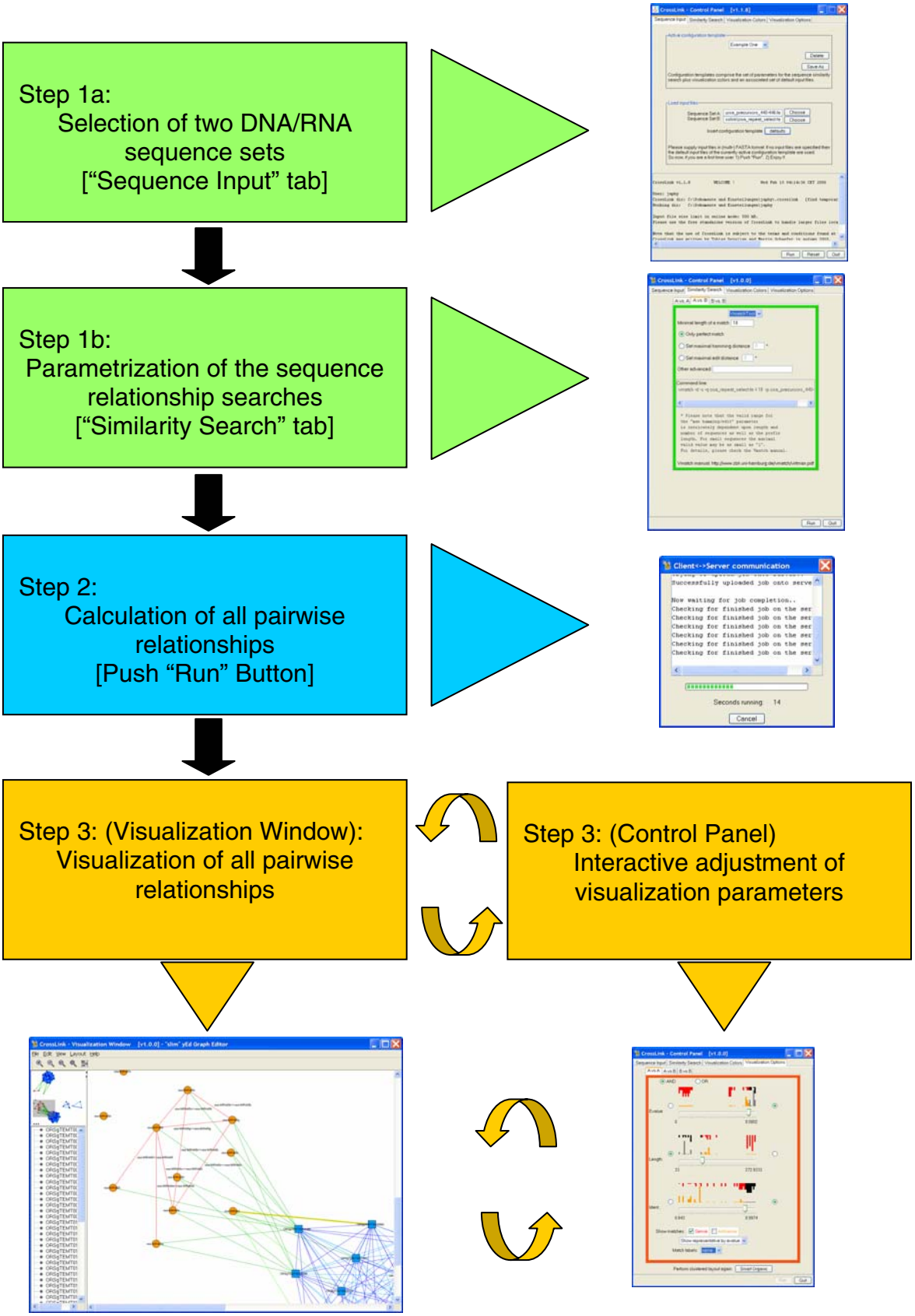


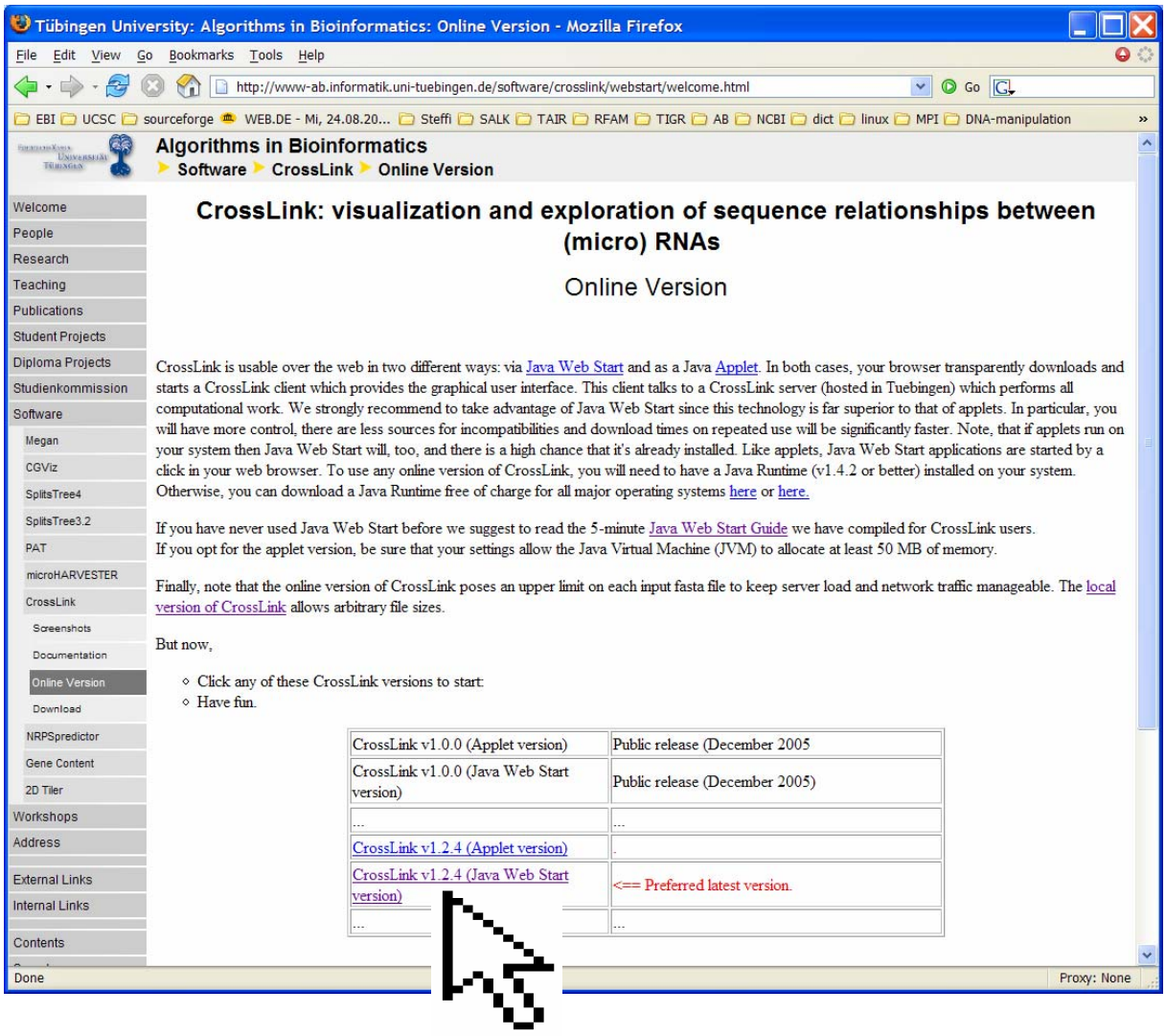
Figure 1 – A simple network with 5 sequences and 8 matches

Usage overview

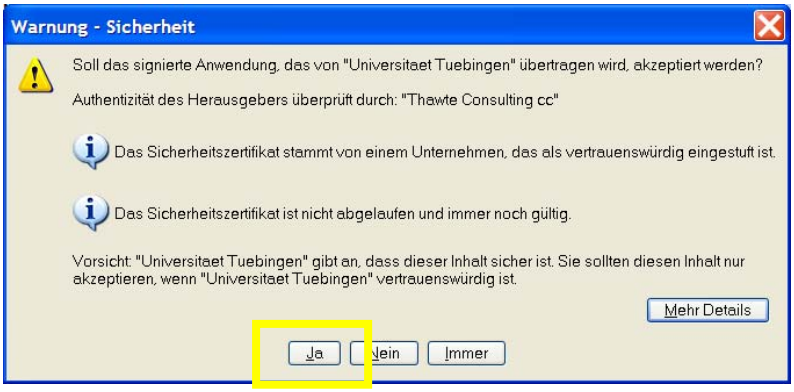
This flow-chart provides a schematic overview of the steps performed during a typical CrossLink session:



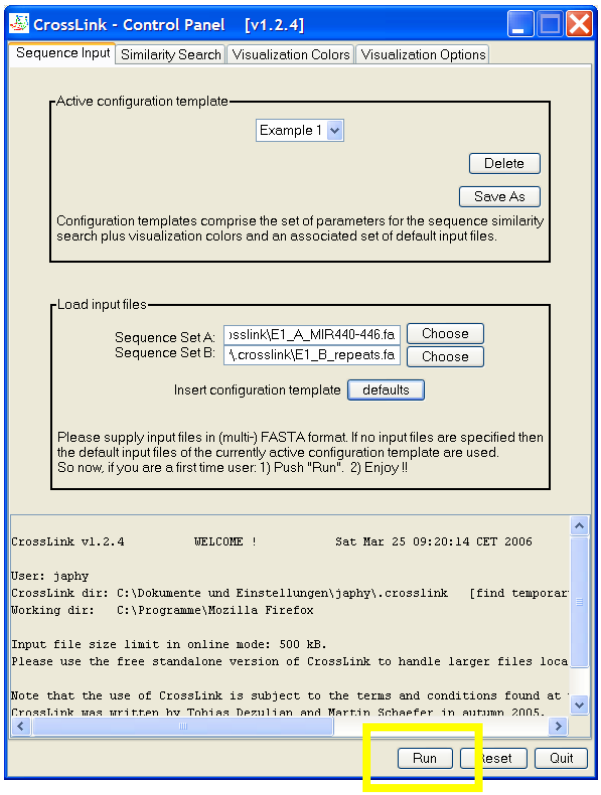
Quick beginner's tour


Tübingen University: Algorithms in Bioinformatics: Online Version - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www-ab.informatik.uni-tuebingen.de/software/crosslink/webstart/welcome.html
Algorithms in Bioinformatics
Software CrossLink Online Version
Welcome
People
Research
Teaching
Publications
Student Projects
Diploma Projects
Studienkommission
Software
Megan
CGViz
SplitsTree4
SplitsTree3.2
PAT
microHARVESTER
CrossLink
Screenshots
Documentation
Online Version
Download
NRPSpredictor
Gene Content
2D Tiler
Workshops
Address
External Links
Internal Links
Contents
CrossLink: visualization and exploration of sequence relationships between (micro) RNAs
Online Version
CrossLink is usable over the web in two different ways: via Java Web Start and as a Java Applet. In both cases, your browser transparently downloads and starts a CrossLink client which provides the graphical user interface. This client talks to a CrossLink server (hosted in Tuebingen) which performs all computational work. We strongly recommend to take advantage of Java Web Start since this technology is far superior to that of applets. In particular, you will have more control, there are less sources for incompatibilities and download times on repeated use will be significantly faster. Note, that if applets run on your system then Java Web Start will, too, and there is a high chance that it's already installed. Like applets, Java Web Start applications are started by a click in your web browser. To use any online version of CrossLink, you will need to have a Java Runtime (v1.4.2 or better) installed on your system. Otherwise, you can download a Java Runtime free of charge for all major operating systems here or here.
If you have never used Java Web Start before we suggest to read the 5-minute Java Web Start Guide we have compiled for CrossLink users.
If you opt for the applet version, be sure that your settings allow the Java Virtual Machine (JVM) to allocate at least 50 MB of memory.
Finally, note that the online version of CrossLink poses an upper limit on each input fasta file to keep server load and network traffic manageable. The local version of CrossLink allows arbitrary file sizes.
But now,
Click any of these CrossLink versions to start:
Have fun.
CrossLink v1.0.0 (Applet version)
Public release (December 2005
CrossLink v1.0.0 (Java Web Start version)
Public release (December 2005)
CrossLink v1.2.4 (Applet version)
CrossLink v1.2.4 (Java Web Start version)
<== Preferred latest version.
Done
Proxy: None

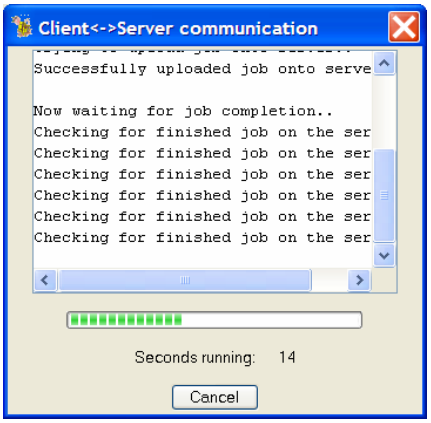
You have Java installed. Click the online Java Web Start version to start.


Warnung - Sicherheit
Soll das signierte Anwendung, das von "Universitaet Tuebingen" übertragen wird, akzeptiert werden?
Authentizität des Herausgebers überprüft durch: "Thawte Consulting cc"
Das Sicherheitszertifikat stammt von einem Unternehmen, das als vertrauenswürdig eingestuft ist.
Das Sicherheitszertifikat ist nicht abgelaufen und immer noch gültig.
Vorsicht: "Universitaet Tuebingen" gibt an, dass dieser Inhalt sicher ist. Sie sollten diesen Inhalt nur akzeptieren, wenn "Universitaet Tuebingen" vertrauenswürdig ist.
Mehr Details
Ja
Nein
Immer

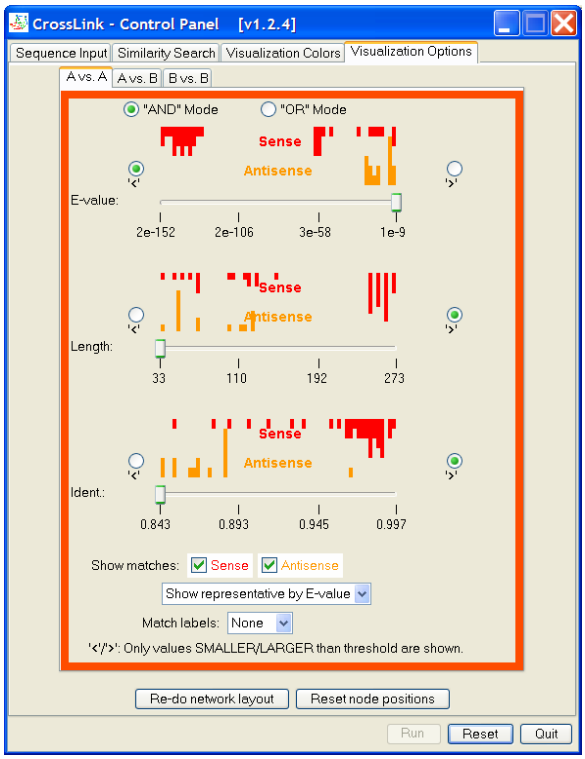
Do you trust a CrossLink application from Tuebingen University? Yep.


CrossLink - Control Panel [v1.2.4]
Sequence Input
Similarity Search
Visualization Colors
Visualization Options
Active configuration template
Example 1
Delete
Save As
Configuration templates comprise the set of parameters for the sequence similarity search plus visualization colors and an associated set of default input files.
Load input files
Sequence Set A:
Sequence Set B:
Choose
Choose
Insert configuration template
defaults
Please supply input files in (multi-) FASTA format. If no input files are specified then the default input files of the currently active configuration template are used.
So now, if you are a first time user: 1) Push "Run". 2) Enjoy !!
CrossLink v1.2.4 WELCOME ! Sat Mar 25 09:20:14 CET 2006
User: japhy
Working dir: C:\Programme\Mozilla Firefox
Input file size limit in online mode: 500 kB.
Please use the free standalone version of CrossLink to handle larger files loca
Run
Quit

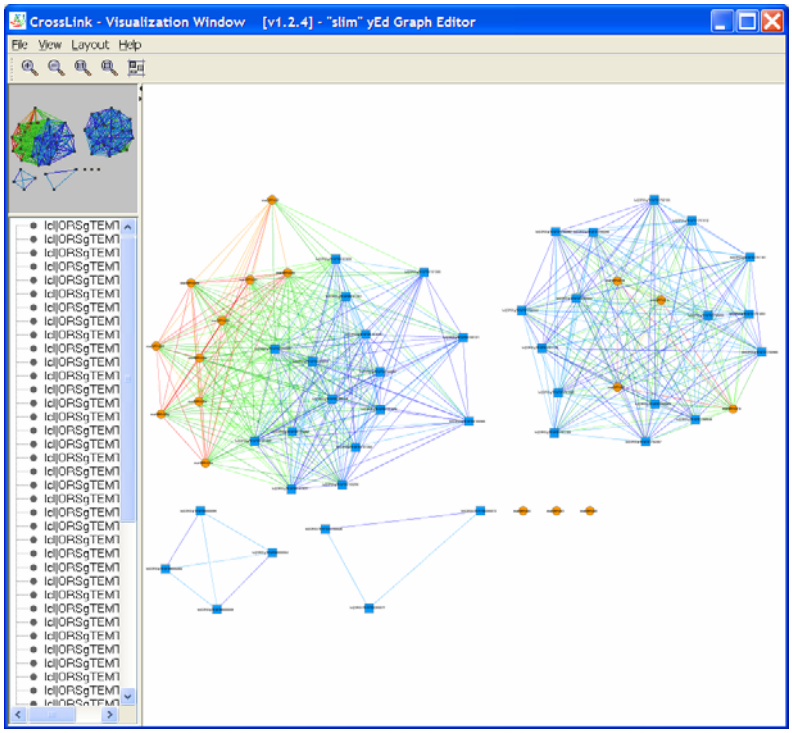
We are happy with “Example 1” for now and click the “Run” button without further ado.


Client<->Server communication
Successfully uploaded job onto serve
Now waiting for job completion..
Checking for finished job on the ser
Checking for finished job on the ser
Checking for finished job on the ser
Checking for finished job on the ser
Checking for finished job on the ser
Checking for finished job on the ser
Seconds running: 14
Cancel

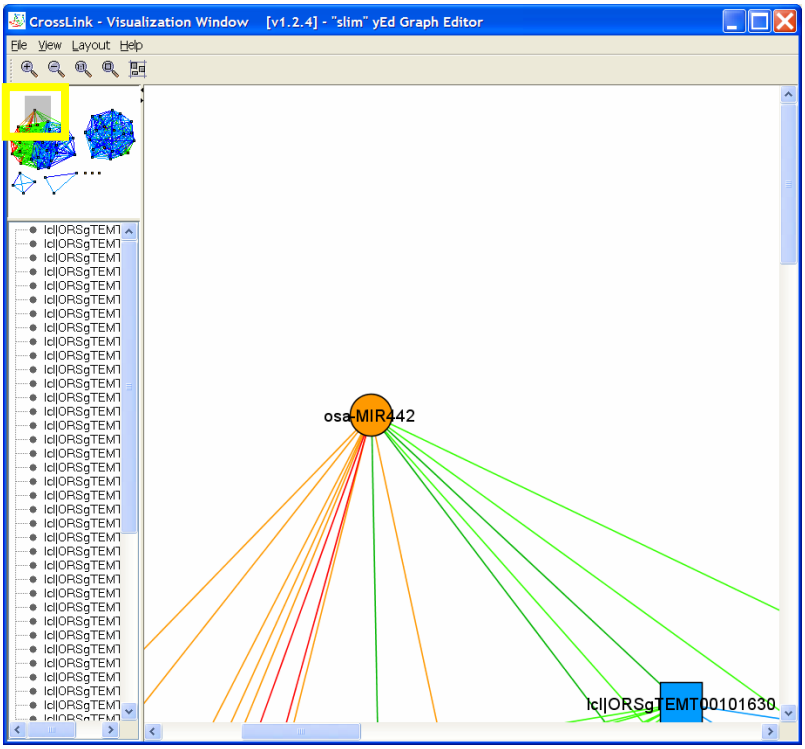
We wait for the server to perform all three relationship searches. For “Example 1” this should take less than a minute.



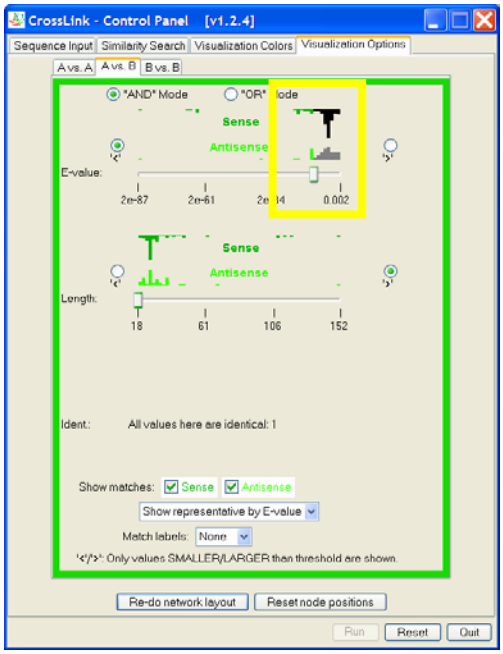
Things seem to have gone well and we get a success screen.



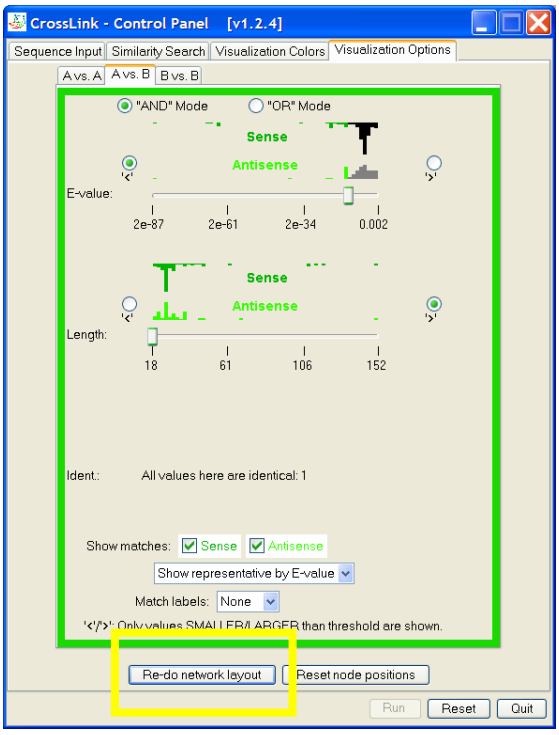
The visualization window opens up in addition. Dive into the graph by zooming with the mouse wheel.



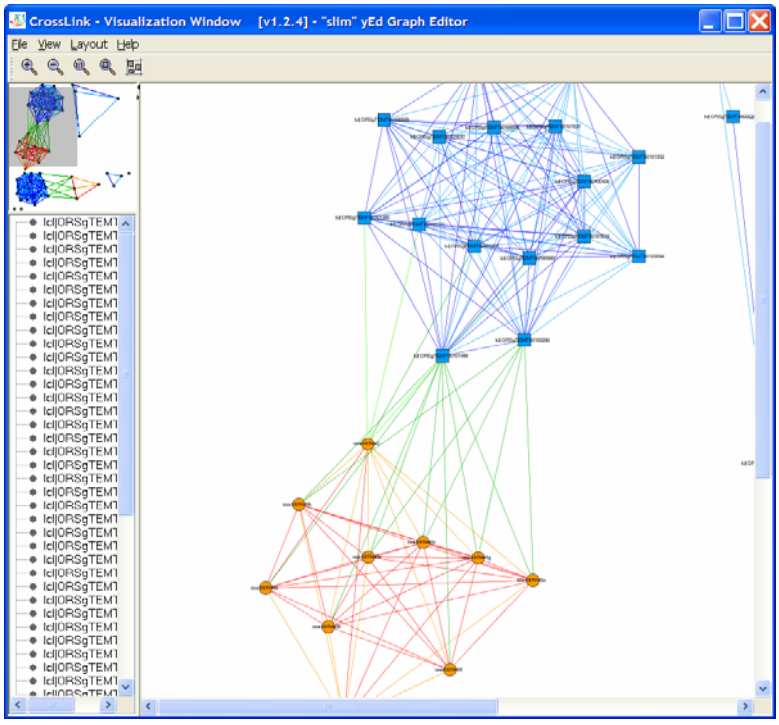
Drag the gray square around the overview panel by pressing and holding the left mouse button. The graph looks a bit messy because too many edges are shown.



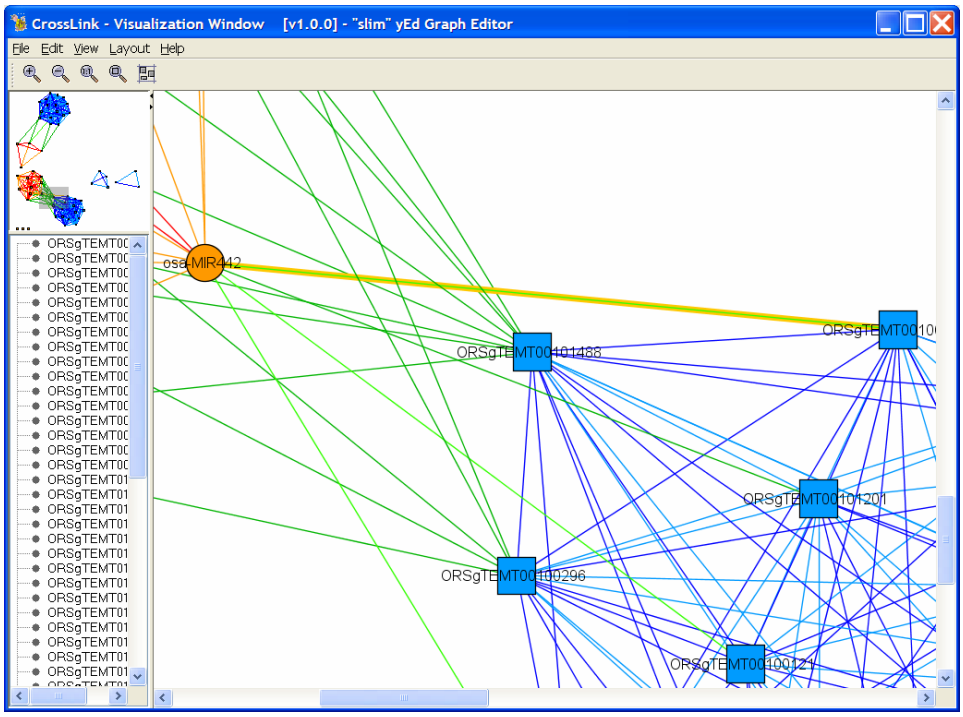
We mask most of the insignificant green edges by moving the “E-value” slider in the “A vs. B” tab of the visualization options panel. This masks all matches above the new evalue threshold.



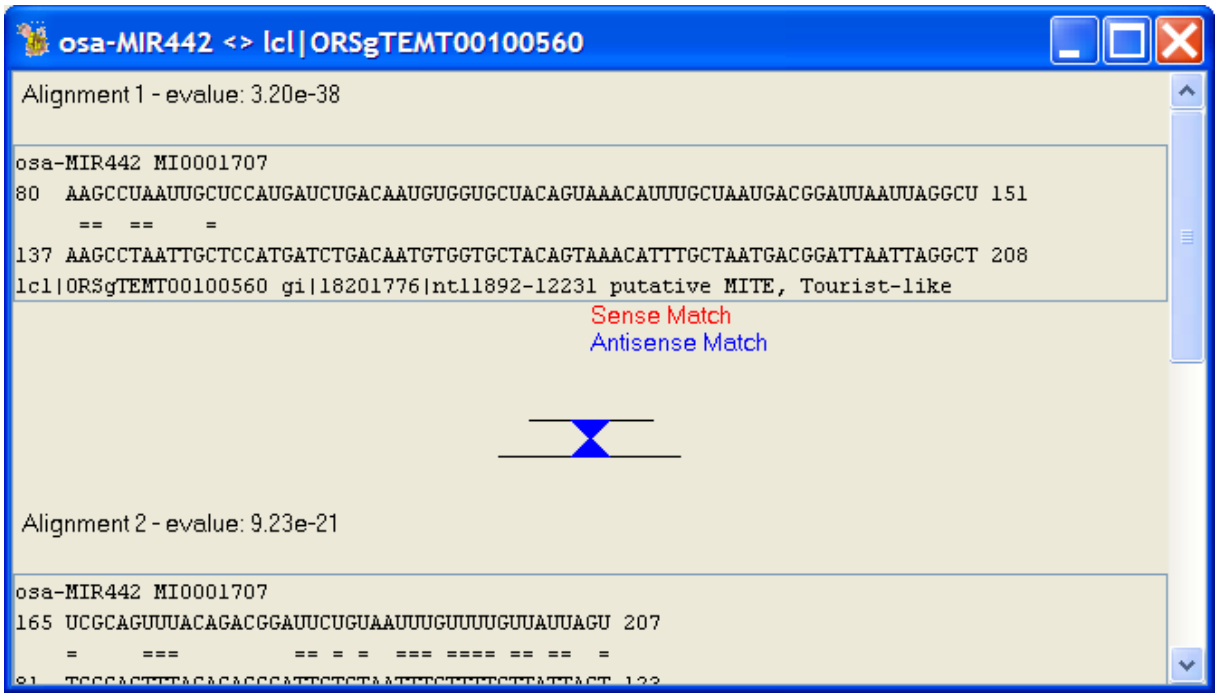
We trigger a re-layouting of the graph with (now) fewer green edges.



And find a much better separation of the “A” set (red) and the “B” set (blue).



Double-clicking on a green edge reveals..



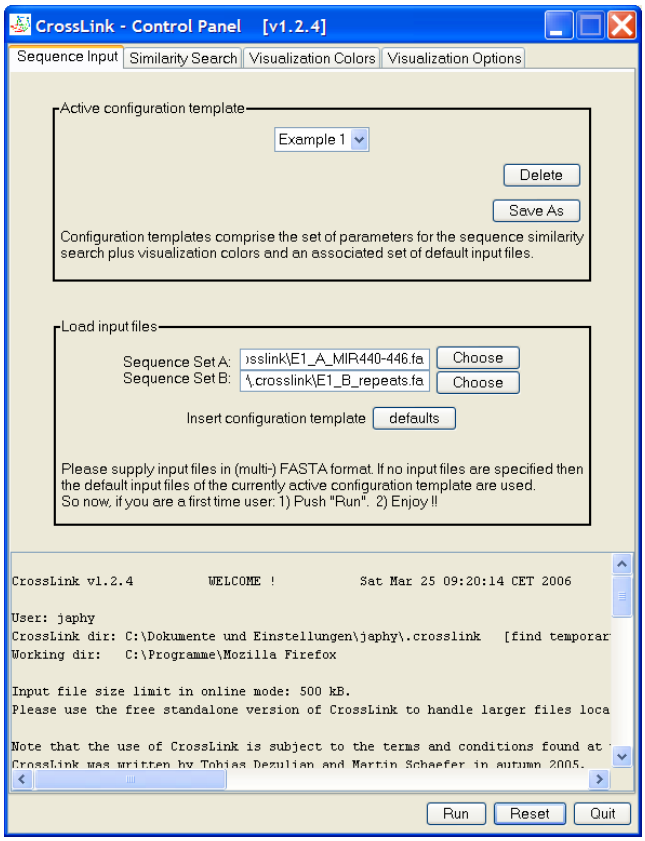
..that the rice microRNA osa-MIR442 seems quite similar to a Tourist-like MITE.

User Interface

CrossLink's user interface essentially consists of two windows: the Control Panel and the Visualization Window.
The Control Panel guides the user through the first two steps as show in the usage overview section. The final exploration phase involves both the Control Panel and the Visualization Panel.

Control Panel

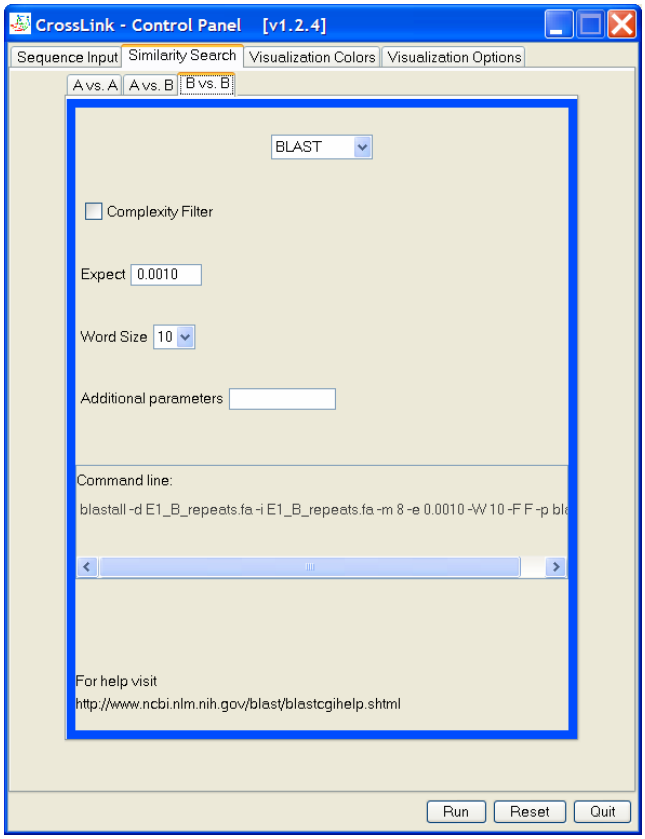
The Control Panel hosts 4 sub-panels. We describe each one in turn now:


CrossLink - Control Panel [v1.2.4]
Sequence Input
Similarity Search
Visualization Colors
Visualization Options
Active configuration template
Example 1
Delete
Save As
Configuration templates comprise the set of parameters for the sequence similarity search plus visualization colors and an associated set of default input files.
Load input files
Sequence Set A:
Sequence Set B:
Choose
Choose
Insert configuration template
defaults
Please supply input files in (multi-) FASTA format. If no input files are specified then the default input files of the currently active configuration template are used.
So now, if you are a first time user: 1) Push "Run". 2) Enjoy !!
CrossLink v1.2.4 WELCOME ! Sat Mar 25 09:20:14 CET 2006
User: japhy
Working dir: C:\Programme\Mozilla Firefox
Input file size limit in online mode: 500 kB.
Please use the free standalone version of CrossLink to handle larger files loca
Note that the use of CrossLink is subject to the terms and conditions found at
Run
Reset
Quit

The Sequence Input panel allows one to select a configuration template. It comprises the set of parameters for each of the relationship searches plus visualization colors and an associated set of default input files. Configuration templates are useful to repeat an exploration task at a later time without having to re-enter all parameters. Also, the whole parameter set of one exploration task can be easily be applied to a different pair of sequence sets.

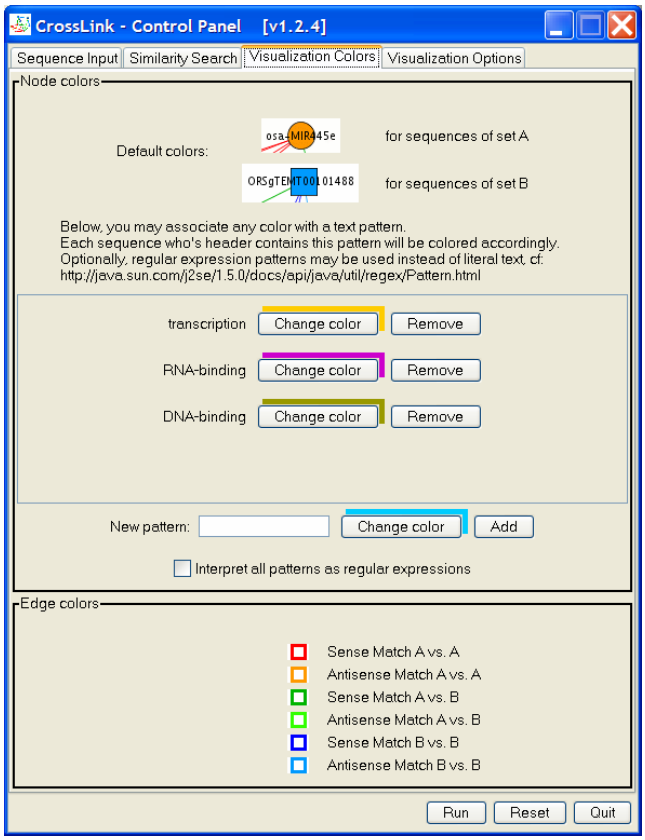
When input files are chosen and "Run" is pushed, the input files given are actually copied to the .crosslink directory, where they are handled. Note that switching to a different configuration template does NOT change the stated input files. You may, however, override the given input files with the default files associated with the template by pressing the "Insert configuration template *defaults*" button.

On the bottom of the Sequence Input panel you find the console which keeps a crude record of the actions performed during this exploration session. In case of errors in the tools used (e.g. when an undefined parameter has been passed to BLAST), the error message returned by the tool will be found here.

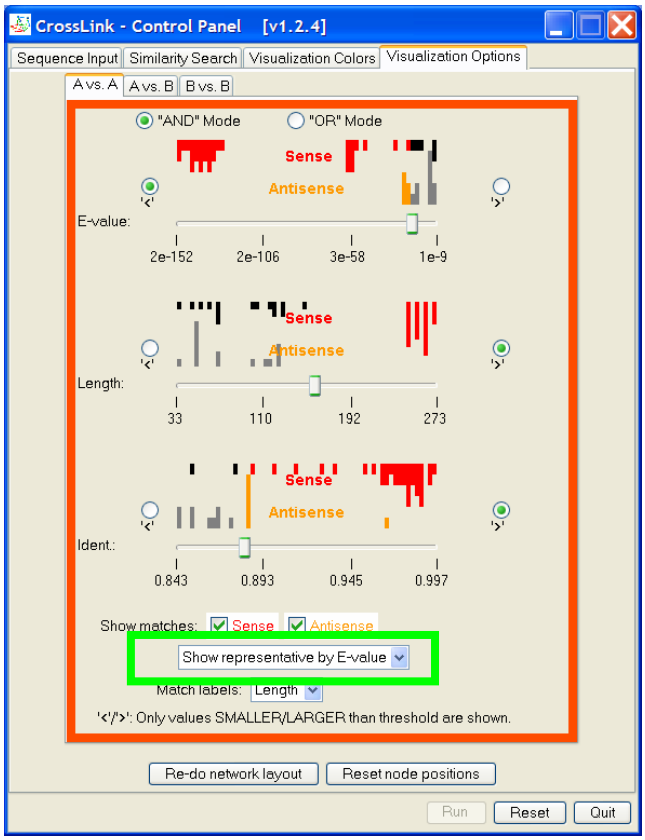


The Similarity Search panel allows selection of a tool for each of the three searches plus parametrization of each tool. Note that you may pass any parameter to each tool which it would accept on the command line if invoked directly. Consequently, you can use the full range of features of each tool. The actual command line which will be passed to each tool is displayed on the bottom of the parametrization panel.

Important remark: BLAST calculates E-values based on the size of the target database, which is constructed from the second set of sequences passed to BLAST. This leads to different E-values for a particular query/target pair when sequences are added or removed to the second sequence set – which some of our test-users found quite confusing ;)



The Visualization Color panel allows the association of colors to patterns of text as described in the article. Note that optionally all patterns stated may be interpreted as regular expressions in the format described here:
http://java.sun.com/j2se/1.5.0/docs/api/java/util/regex/Pattern.html
Edge colors are not changeable. An edge is colored according to the match set it belongs to.

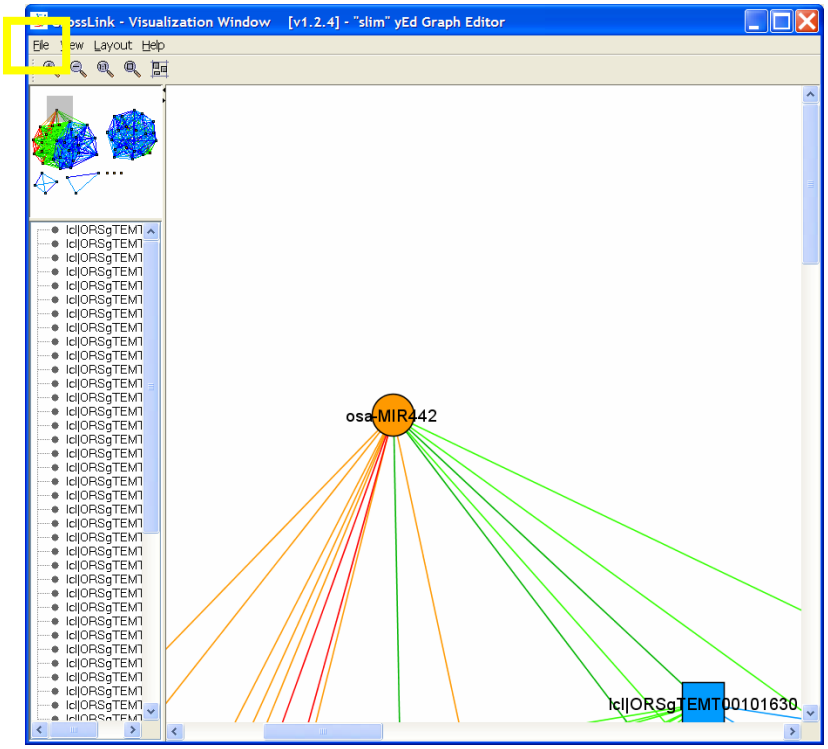


The Visualization Options panel allows interactive manipulation of the network shown in the Visualization Window. Dragging the slider of histogram marks some edges in gray, which are consequently suppressed from visualization. Using the radio buttons to the left and right of each histogram allows the threshold to apply to matches which are smaller or larger, respectively. Using the “AND” mode only visualizes those matches which pass each of the three thresholds. The “OR” mode visualizes those matches which pass at least one of the given thresholds.
Using the checkboxes labeled “Show matches: Sense/Antisense” one may suppress all sense and/or antisense matches altogether. A dropdown box (marked in green above) allows one to switch from “Single match mode” (equivalent to “Show all matches”) to “Multiple match mode” (equivalent to “Show representative”) and, if the case of “Show representative” allows one to select a criterion to select one representative match from each match set. This representative match is selected depending on one of the following selections:

- Show representative by E-value
  The match with the smallest E-value represents this match set
- Show representative by Length
  The match with the largest length represents this match set
- Show representative by Identity
  The match with the largest fraction of identical nucleotides within the alignment represents this match set

In “multiple match mode”, all matches between one pair of sequence (nodes) are represented by one link in the network. The attibute shown when “Match labels” is selected is the attribute of the representative match of the pertaining match set.

Visualization Window



Most menu things here should be self-explanatory.
One remark, however:

- ►File►Save allows you to save the network in YGF format. You can subsequently use the free yEd graph editor from yWorks (www.yworks.com), available at http://www.yworks.com/en/products_yed_about.htm (web start version, applet version, standalone version) to load and modify the saved network in a large variety of ways for printing.

Example templates

CrossLink provides three example configuration templates along with the corresponding sequence files. To try out CrossLink, one merely has to select one of the examples and press the "Run" button. When trying the examples in sequence be sure to load the default input files associated with each example – as switching the configuration template does not change the current input files (but provides new associated default input files that may be loaded by pushing the "Insert configuration template *defaults*" button).
The following example scenarios are provided:

- Example 1: Sequence set A consists of all rice microRNAs of families 440-446 available from miRBase. Sequence set B contains a subset of repetitive rice sequences downloaded from the TIGR Rice Genome Annotation Database. It is immediately visible that e.g. the rice microRNA family 445 exhibits very close sequence similarity to a family of repetitive rice sequences. Initially displaying a multitude of links in a tangle, this example demonstrates the

power of the interactive histograms to focus on relevant relationships.

We have compiled this example after an exploration of all rice microRNA families had suggested that at least some rice microRNA sequences seem to be associated with repetitive elements.The given subset of repetitive rice sequences has been quite randomly picked from the much larger set of microRNA-related sequences available at TIGR.

- Example 2: Sequence set A consists of all Arabidopsis microRNA precursors available at miRBase. Sequence set B contains all (~2000) sequences contained in the Arabidopsis Small RNA Project Database to date. Setting these two sets in relation to each other allows one to assess which microRNA families have been sequenced by the ASRP project. This example also demonstrates CrossLink's ability to handle large sets of sequences and also shows the power of the spring-embedding algorithm in clustering microRNAs into families.

  The ASRP sequence set should contain microRNAs in a proportion roughly equivalent to their abundance in cellular RNA. Thus is seems informative that some microRNA families cluster with very many ASRP sequences and others only with very few.

- Example 3: Sequence set A consists of the Drosophila microRNAs dme-miR-3, dme-miR-4 and dme-miR-5. Sequence set B contains all corresponding targets which have been predicted (with an E-value smaller than one) in a study by Rehmsmeier et al., plus some randomly picked sequences from the same study that have not been predicted as potential targets of these microRNAs. This example demonstrates the use of RNAhybrid, for example revealing that one sequence (accession CG15125) is simultaneously targeted by two different microRNAs. Furthermore, the capability of custom pattern-color associations is shown as each predicted target set of the Rehmsmeier et al. study is associated with its own color (yellow, magenta and cyan for the targets of dme-miR-3, dme-miR-4 and dme-miR-5, respectively) and the non-targets are shown in blue.

  This example uses identical parameters for the RNAhybrid search as given in the Rehmsmeier et al. study: Reported duplexes are forced to form perfect helices from nt 2 to 7 in the microRNAs (using the "-f" parameter). Since we have not been able to unambiguously identify the exact sequences used in the abovementioned study from their given supplementary material, we kindly asked for and received their original material. The initially surprising fact that some sequences (e.g. accession CG13906 and CG3800) bind not only to the microRNA predicted by their study but also to another microRNA of a different family that had not been predicted by them presumeably derives from the fact that we did not calibrate our search separately for each family as they did.

Terms and Conditions

*_LICENSE AGREEMENT_*

This software is being provided to you, the LICENSEE, by the
LICENSOR (Department for Algorithms in Bioinformatics, Tuebingen University)
under the following license. By obtaining and/or copying
this software, you agree that you have read, understood and
will comply with the following terms and conditions.

Permission to use and modify for INTERNAL NONCOMMERCIAL
RESEARCH PURPOSES ONLY, this software and its documentation
is hereby granted, provided that you agree

1) to comply with the following copyright notice and
statements, including the disclaimer, and that the same
appear on ALL copies of the software and documentation,
including duplications and modifications that you make for
internal use: “Copyright 2005 by Department for Algorithms
in Bioinformatics, Tuebingen University. All rights reserved.”

2) NOT TO REENGINEER, DECOMPILE, OR ATTEMPT TO REPRODUCE ANY
COMPONENT NOT PROVIDED IN/AS SOURCE CODE.

THIS SOFTWARE IS PROVIDED “AS IS”, AND THE LICENSOR MAKES NO
PRESENTATIONS OR WARRANTIES, EXPRESS OR IMPLIED.

BY WAY OF EXAMPLE, BUT NOT LIMITATION, THE LICENSOR MAKES NO
REPRESENTATIONS OR WARRANTIES OF MERCHANTABILITY OR FITNESS
FOR ANY PARTICULAR PURPOSE OR THAT THE USE OF THE LICENSED
INFORMATION OR DOCUMENTATION WILL NOT INFRINGE ANY THIRD
PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS OR
THAT THE OPERATION OF THE SITE AT WHICH THIS INFORMATION IS
FOUND WILL BE ERROR-FREE. IN NO EVENT SHALL THE LICENSOR
HAVE ANY LIABILITY TO THE USER FOR CONSEQUENTIAL, SPECIAL,
INCIDENTAL OR OTHER INDIRECT DAMAGES.

The copyright in this software and any associated
documentation shall at all times remain with the LICENSOR
(or any other licensors, which have granted a license to
LICENSOR) and the LICENSEE agrees to preserve the same.